DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

Evolution of an Application Profile: Advancing Metadata Best Practices through the Dryad Data Repository

Edward M. Krause
Metadata Research Center
Drexel University, USA
edward.krause@drexel.edu

Erin Clary
Metadata Research Center
Drexel University, USA
eclary@datadryad.org

Adrian Ogletree
Metadata Research Center
Drexel University, USA
aogletree@drexel.edu

Jane Greenberg
Metadata Research Center
Drexel University, USA
janeg@drexel.edu

Abstract

Dryad is a general-purpose curated repository for data underlying scholarly publications. Dryad's metadata framework is supported by a Dublin Core Application Profile (DCAP, hereafter referred to as application profile). This paper examines the evolution of Dryad's application profile, which has been revised over time, in an operational system, serving day-to-day needs of stakeholders. We model the relationships between data packages and data files over time, from its initial implementation in 2007 to its current practice, version 3.2, and present a crosswalk analysis. Results covering versions 1.0 to 3.0 show an increase in the number of metadata elements used to describe Dryad's data objects in Dryad. Results also confirm that Version 3.0, which envisioned separate metadata element sets for data package, data files, and publication metadata, was never fully realized due to constraints in Dryad system architecture. Version 3.1 subsequently reduced the number of metadata elements captured by recombining the publication and data package element sets. This paper documents a real world application profile implemented in an operational system, noting practical system and infrastructure constraints. Finally, the analysis presented informs an ongoing effort to update the application profile to support Dryad's diverse and expanding community of stakeholders.

Keywords: metadata; metadata schema; application profile; DCAP; Dryad.

1. Introduction

Dryad has been supported by a metadata application profile from its launch in 2007 through the present day (Dryad Data, 2015). An application profile "consist[s] of data elements drawn from one or more namespace schemas combined together by implementors and optimised for a particular local application" (Heery & Patel, 2000). A data element refers to a metadata field, and a namespace schema, or a metadata schema, is a set of standardized metadata elements.

The application profile approach was endorsed by Dryad team members from the beginning, given the need for Dryad metadata to interoperate with other data efforts, and given the desire of Dryad's metadata R&D team to align with semantic web developments and to keep current with metadata developments. Application profiles promote data sharing, interoperability, and linked data, which are all central to the overarching mission of Dryad.

Dryad has been operational since 2008, and has grown at a fairly rapid pace, expanding to accommodate more disciplines and stakeholder organizations. This growth has had an impact on Dryad's functional requirement and day-to-day workflows, expanding the menu of options. These changes have had a significant impact on Dryad's metadata application profile. In this paper, we perform a crosswalk analysis, present domain models, and evaluate individual metadata elements and refinements that have changed over time. The paper also serves to document the change in an

DC PAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

application profile over time and to produce an updated representation of Dryad's current metadata practice.

1.1. What is Dryad?

Dryad is a curated digital repository for data underlying peer-reviewed scholarly literature. The stated mission of the repository is to "make the data underlying scholarly publications discoverable, accessible, understandable, freely reusable, and citable" (Dryad, 2015). Dryad is also committed to the long-term preservation of archived data (Mannheimer et al., 2014). While Dryad began as an infrastructure for data archiving in evolutionary biology and ecology, the scope of the repository has since expanded. Dryad has developed into a general-purpose repository for long-tail scientific data, and the repository currently accepts data from a wide variety of disciplines, including medical and social sciences.

Each data package in Dryad is linked to its associated publication, and Dryad stores metadata related to the data package and its files, in addition to metadata derived from the publication. Dryad works with a data package model, in which a data package can have one or more data files. Dryad's chief mission is to make data discoverable and reusable for scientific endeavors. Metadata is essential for these steps, and for fulfilling Dryad's mission.

1.2. Dryad's Early Application Profile Work (2007-2009)

Since its origins, Dryad has actively incorporated the Dublin Core Abstract Model (DCMI, 2007), adhering to the Singapore Framework for Dublin Core Application Profiles (DCMI, 2008), into a metadata best practice (Powell et al., 2007; Nilsson et al., 2008; Greenberg et al., 2009). These two abstract information models, developed by the Dublin Core community, represent efforts to move from the resource-driven legacy approach representing an information package toward focusing on the component parts of a resource description. The initial goals of developing an application profile for Dryad were twofold; an immediate short-term concern was to make content available in DSpace through an XML schema, and in the long-term, to align with the Semantic Web (Greenberg et al., 2009).

The first version of Dryad's application profile (v1.0) was developed in 2007, before the release of the Singapore Framework guidelines. Although the Singapore Framework had not yet been published, development of Dryad's metadata application profile still began with the critical first steps of defining the repository's functional requirements and creating a domain model, as prescribed in the Guidelines for Dublin Core Application Profiles (Coyle & Baker, 2009). These first steps are reported on in more detail in Dube et al. (2007) and White et al. (2008).

TABLE 1: Dryad DCAP v.3.1: Metadata elements (Dryad, 2013).

Data Package	Data File
dcterms:type	dcterms:type
dcterms:creator	dcterms:creator
dcterms:dateSubmitted	dcterms:dateSubmitted
dcterms:available	dcterms:available
dcterms:title	dcterms:title
dcterms:identifier	dcterms:identifier
dcterms:description	dcterms:description
dcterms:subject	dcterms:subject
dwc:scientificName	dwc:scientificName
dcterms:spatial	dcterms:spatial
dcterms:temporal	dcterms:temporal
external	embargoedUntil
dcterms:references	dcterms:rights
bibo:pmid	dcterms:format
bibo:Journal	dcterms:provenance
dcterms:hasPart	dcterms:isPartOf

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

1.3. Dryad, DSpace, and Further Application Profile Development (2013-2015)

The most current version of the application profile is v3.1, published in an XSD file (Dryad, 2013). The metadata elements included in v3.1 are listed in Table 1. Elements that are shaded green are the ones that are used to describe both the data package and file. These elements are intended to document bibliographic metadata of the associated publication, scope and coverage of the data files, and the relationship between the data files, the data package, and the publication, each of which is represented by a unique identifier. The profile includes elements from several namespaces, including Dublin Core (DCMI, 2012), Darwin Core (Darwin, 2015), and Dryad's own namespace. Dryad has been implemented on version 1.8 of the DSpace framework (DSpace, 2015). While the latest version of DSpace released, as of this publication, is version 5.0, Dryad has not upgraded to a later version of the framework due to the risk of unforeseen upgrade incompatibilities with the extensive customizations of the system architecture made by Dryad developers. Though Dublin Core does not support dot-notation for representing metadata elements and the associated refinements (e.g. dcterms:coverage.spatial), DSpace continues to use this type of notation internally to represent metadata elements. During automated metadata harvesting, internal metadata elements are converted to Dublin Core compliant properties from the *terms* namespace.

Dryad is built on an early version of DSpace and elements are stored internally. DSpace is among one of the most popular repository software used for digital libraries, storage of offprints, and other digital creative outputs of an institution. Among several well-known DSpace users are Cornell University Libraries, Deep Blue at the University of Michigan, and Rice University's TIMEA digital archive. DSpace was selected for Dryad because of its open source status, its user-friendly interface for scientists/researchers as depositors, and because it could be installed out of the box. Dryad has worked with Atmire (http://atmire.com/website/) since the beginning to better accommodate scientific data deposits.

Ongoing development of an operational system, with real users and day-to-day needs, has been an exciting undertaking for the Dryad team. The progress has been consistent, keeping Dryad fully functional, although, as one may anticipate, there have been delays in keeping pace with the most current DSpace release, particularly given the unique nature of Dryad. Another important point is that DSpace provides access to an extensive list of Dublin Core metadata properties along with properties from additional namespaces within the curation module; however, the current metadata infrastructure doesn't fully align with the DCMI's DCAP for rendering RDF metadata, and the syntactic encoding differs. Metadata generated via DSpace can be converted to RDF, although this has not been a chief priority for Dryad at this time, with current day-to-day, real-world needs servicing clients and making descriptions accessible. The aim of being fully compliant with DCMI, aligning with the Singapore Framework, and the DCAM (Dublin Core Abstract Model) is part of Dryad's two-pronged approach, and has been documented in Greenberg et al., (2009). This paper presents an account of the activity that is impacting the day-to-day work, and the guiding research objectives are outlined in the next section.

2. Research Objectives

This study is the first step in a larger process to document and assess Dryad's metadata application profile. Dryad's initial metadata scheme was devised to allow for data ingest, and to support preservation, access, and basic usage of data (Dube, 2007). Dube et al. (2007) also proposed long-term goals for the metadata scheme, including expanded support for data use, extended interoperability and support for semantic web functionalities.

Dryad's initial disciplinary focus was evolutionary biology. Today, the repository is still heavily in the bioscience area, although Dryad is promoted as a general-purpose repository, and there is a growing representation from a wide array of disciplinary fields, ranging from the biomedical field to physics, chemistry, information science, and social sciences. This change, and stakeholder growth (including more publishers and organizations) has resulted in new functional requirement, which in turn have had an impact on the application profile. Given the pace of

⊛DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

change, it seems timely to revisit the application profile work and document the current practice. The goal of the research reported on in this paper is to examine how Dryad's application profile has evolved from its first inception in 2007 as version 1.0 through the last update in 2013 as version 3.1. This study will document changes in the element set over time. An end goal of this study is to align Dryad's application profile with current practice as version 3.2 and to propose next steps to update the application profile. This will help Dryad to maintain high quality metadata practice, and help provide a platform for attaining higher-level objectives of automatic data synthesis as described in 2007.

3. Methods

To investigate the goals and methods outlined in Section 2, we used a crosswalk analysis to compare each version of the application profile and modeled the relationship between data package, data file and publication that was represented by each application profile. While crosswalk analyses are primarily used to facilitate interoperability among applications that may use different metadata schemas by mapping metadata elements, semantics, and syntax from each schema to determine their compatibility (NISO, 2004), we conducted a modified crosswalk analysis to examine changes in metadata usage across the different versions of Dryad's application profile. Domain models define the basic structures and relationships of digital entities (Nilsson et al., 2009). In Dryad, each entity - data package, data file and publication - is described by a set of metadata elements. Changes in the domain models across application profile versions reflect changes identified in the crosswalk analysis. Each version of the application profile was compared to the previous iteration, and changes in element usage were documented. Lastly, an updated version of the application profile, version 3.2, was created to report on current metadata practices in Dryad.

4. Results and Discussion

The results and contextual discussion that follow detail the crosswalk analysis, Dryad's changing domain models, and version changes.

4.1. Crosswalk Analysis

The Dryad application profile has drawn from multiple metadata schemas throughout its version history. The current profile includes elements from Dublin Core (namespace: dcterms), Darwin Core (namespace: dwc), and Publishing Requirements for Industry Standard Metadata (namespace: prism) (Idealliance, 2015). The application profile also includes Dryad namespace elements, which represent concepts required for repository functionality that were not found in other schemas. For instance, Dryad captures the number of page views and downloads of each data file with the elements dryad:pageviews and dryad:downloads. As mentioned earlier, DSpace uses a dot-notation to express elements and their refinements internally, and this is how some metadata elements will be described in the Results and Discussion. Table 2 explains the relationship between Dryad/DSpace internal elements and their corresponding external notations as they are represented in automated metadata harvests.

Early versions of the application profile included elements from Data Documentation Initiative (namespace: DDI) (DDI, 2009), Journal Publishing Tag Set (namespace: journalpublishing3) (NCBI, 2012), Preservation Metadata: Implementation Strategies (namespace: PREMIS) (LoC, 2015), and Bibliographic Ontology Specification (namespace: bibo) (Bibliographic, 2009); however, elements from these schemas are not currently used. Many of the metadata elements from the discontinued schemas are now represented as Dublin Core refinements. For instance, version 2.0 used elements from the PRISM and Journal Publishing Tag Set schemas to store publication citation metadata, while version 3.0 replaced and expanded upon the PRISM concepts with elements from the Bibliographic Ontology Specification. In versions 3.1 and 3.2, the elements used to store citation information were collapsed into a single field, dcterms:identifier.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

The crosswalk analysis revealed four possible cases for each metadata element in the application profile: 1) The element and the concept it represents (an element-concept pair) did not change, and is present in all iterations of the application profile. 2) The concept did not change, but the element that was used to represent that concept did change from version to version. 3) Elements and concepts are added, and 4) Elements and concepts are phased out.

Metadata elements that are used in each version of the application profile include those that represent descriptive, spatial and temporal characteristics, digital identifiers, types, relationships, subjects, and taxonomic classification. Other metadata concepts have remained constant through each version of the application profile, but are represented by different metadata elements over time. For instance, the embargo end date, which is the date on which a data file will be made available for download, was initially recorded at dcterms:available. This concept was later represented by the element dcterms:embargoedUntil, while dcterms:available was repurposed to represent the date and time a curator approved a data package into the archive. This definition of dcterms:available was more congruent with the Dublin Core definition of this term as a "date (often a range) that the resource became or will become available" (DCMI, 2012). However, the metadata describing a data file may be made available at the public website before the file itself is available for download, hence the embargo date refinement for data files within a data package.

Each version of the application profile is a snapshot of Dryad's workflow and functionality at a particular point in time. While many of the elements of versions 1.0 and 2.0 were phased out prior to the current version, version 3.0 introduced multiple concepts and elements that are currently used; these element-concept pairs chronicle the evolution of repository functionality. For instance, an element to record provenance metadata, dcterms:description.provenance, was added in version 3.0. Metadata, including date, time and name of the person who performed an action, are automatically captured at ingest, and each time a data package changes workflow stages. The crosswalk analysis also depicts a more recent increase in the number of concepts and elements added to the application profile in version 3.2. For instance, publication blackout dates allow for automated release of submissions to the archive, correlating to the expected release of the article online by the publisher. Recent element changes demonstrate an increase in advanced functions, including automation of certain curation tasks.

4.2. Dryad's Changing Domain Models

Comparison of the domain model versions (Figure 1) provides additional context to the application profile version changes.
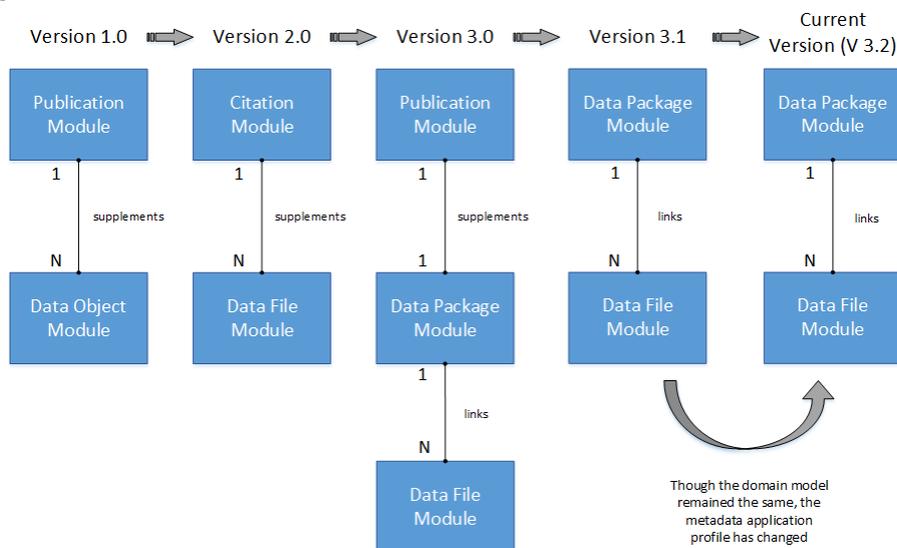

FIG. 1: Dryad domain model versions.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

Versions 1.0 and 2.0 use a similar domain model, with a single publication or citation module linked to multiple data objects or files. The publication/citation modules contain metadata pertaining to the publication associated with the archived datasets as well as metadata that links to the data file modules. In versions 1.0 and 2.0, one or more data files could be associated with one and only one publication or citation. Elements pertaining to the associated journal citation were added in version 2.0, including article title, journal name, volume and issue, which increased the granularity with which journal citations were captured in Dryad.

The evolution from version 2.0 to 3.0 of the application profile domain model shows an expanded set of entities, where the publication module is split from the data package module and the data package module is linked to the data file module. Additional journal citation elements were added in version 3.0, increasing the granularity of the journal citation concept. The additional journal citation elements include ISSN/EISSN, PMID, and status. At this point in the application profile development, provenance metadata was also included to track workflow step changes and the users who perform the workflow changes. Additional metadata elements were required to identify and link the three entities represented in version 3.0. Version 3.0 specifies a 1:1 relationship between the publication and data package module and a 1:N relationship to the data file module. This was an effort to bring back the publication as a first-class object within Dryad. It is important to note that version 3.0 was an idealized version of the Dryad application profile, and was never fully implemented due to constraints on the Dryad system architecture. In addition, it was determined to have few practical benefits to Dryad's users.

When Dryad was initially developed, there was no concept of the data package; instead, the domain model only included publications and associated data files. As Dryad grew, the idea of a data "package" was introduced. The records in Dryad that were formerly used to represent publications were changed to be data packages, though they still contain some information related to the publication. By recombining the publication and data package modules, version 3.1 represents a more feasible, scaled-down version 3.0, while still retaining the 1:N relationship between the package and file modules. With only two domain model entities in this version, fewer identifiers and relational elements were required to describe the contents of and relationships among the entities. Version 3.1 also demonstrated a consolidation of metadata elements related to the associated journal publication into a single citation metadata element.

As noted in Figure 1, version 3.2 of the application profile preserves the domain model of version 3.1, but includes changes in the metadata elements it represents. V3.2 includes elements for the manuscript number of the associated publication and a publication blackout release date, which corresponds to the date the associated publication will be released online.

4.3 Dryad Application Profile Version 3.2

The updated Dryad application profile is presented in Appendix A of this article and also published in Dryad (Krause et al., 2015). An example of three metadata elements is presented in Table 2. This table documents the namespace and name of the element as it is represented internally by DSpace; the element as it is represented externally as metadata is harvested by an API, a URI, a definition; the module in which the element is included, the obligation, and cardinality. Elements may be located in the data package module the data file module or both modules. The data package module contains 24 metadata elements from the Dublin Core, Darwin Core, and PRISM schemas, as well as from the Dryad namespace. Many of these elements, such as spatial coverage, subject, and scientific name, can be automatically propagated to the data file module. This reduces the effort required for the submitter to provide richer metadata at the individual file level. While the most common Dryad workflow is archiving data as part of the publication process, the repository is now supporting inclusion of data in the peer review process for several journals. This new workflow has had an impact on the set of metadata elements implemented by Dryad. For example, the metadata element dcterms:manuscriptNumber links a manuscript to its associated data package, allowing publishers to consider the associated datasets that underlie submitted manuscripts before they are published. The updated data file module

◉ DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

contains 21 metadata elements from the Dublin Core and Darwin Core schemas and the Dryad namespace. Data files are linked to the data package module through the dcterms:ispartof and dcterms:relationhaspart metadata elements, which point to the digital object identifier (DOI) of the linked modules.

TABLE 2: Selected Dryad Metadata Application Profile Elements, Version 3.2.

Internal Element Representation (DSpace):					dcterms:contributor.author	
External Element Representation (Metadata Harvesting APIs):					dcterms:creator	
URI:		http://purl.org/dc/terms/creator				
Definition:		Authors on publication / Authors of data submission				
Module(s):	Package & File	Obligation:	Required	Cardinality:		Repeatable
Internal Element Representation (DSpace):					dcterms:coverage.spatial	
External Element Representation (Metadata Harvesting APIs):					dcterms:spatial	
URI:		http://purl.org/dc/terms/spatial				
Definition:		Spatial description of the data specified by a geographic description and/or geographic coordinates				
Module(s):	Package & File	Obligation:	Optional	Cardinality:		Repeatable
Internal Element Representation (DSpace):					dcterms:coverage.temporal	
External Element Representation (Metadata Harvesting APIs):					dcterms:temporal	
URI:		http://purl.org/dc/terms/temporal				
Definition:		Temporal description of the data, as geologic timespan or dates of data collection/research				
Module(s):	Package & File	Obligation:	Optional	Cardinality:		Repeatable

5. Conclusion

This paper reports on efforts to align Dryad's application profile with current practice, and will be published as Version 3.2. Application profiles promote data sharing, interoperability, and linked data, which are all central to the overarching mission of Dryad. We performed a crosswalk analysis and diagrammed domain models to document and compare changes in the application profile. Over time, Dryad has changed the way it conceptualizes the relationships between data files, data packages, and publications. Furthermore, previous work on updating the application profile has revealed limitations in DSpace. Finally, examining which metadata elements and refinements have been added or deleted gives insight to which fields are the most crucial for archiving, preserving, and re-using data.

The data collected in this work is essential in outlining new goals for Dryad's metadata schema. Dryad's community has substantially expanded since its inception in 2007. In addition, the landscape of data repositories and archives has grown a great deal over past decade. New requirements for researchers regarding data deposition should be taken into consideration when deciding what information is collected from researchers about their data. The data collected through this effort will help inform future directions for metadata best practices across scientific data repositories.

As a next step, one of our goals is to publicly declare the Dryad-specific subproperties using the Dryad PURL domain. As indicated above, this paper reports on Dryad's work in day-to-day operational systems, but we have a long term goal to be more fully compliant with the DCMI and align with the Singapore Framework and the DCAM. This much longer-term goal will allow us map our labels onto RDF properties in order to achieve RDF Linked Data interoperability. In addition, we will perform a content analysis and examine a selected set of metadata schemas and

DC PAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

elements, such as DDI or PREMIS. In order to re-evaluate Dryad's functional requirements, it will be necessary to identify and consider new stakeholders (including journals, societies, researchers as both data depositors and data users, funders, and educators) and more complicated curation workflows. In order to determine users' needs, a next step could be to survey different types of users and follow up with more qualitative interviews. In addition, we will need to consider the increasingly diverse data formats and types that are used in the scientific domains represented in Dryad. New metadata elements may be needed to properly describe and preserve clinical data, social science data, and any other scientific data that Dryad could accept in the future. Finally, we will develop concrete objectives for implementing Dryad's metadata best practices, based on a deeper understanding of user needs and limitations of the repository.

Acknowledgements

References

Bibliographic Ontology. (2009). Bibliographic Ontology Specification. Retrieved from http://bibliontology.com/

Carrier, Sarah. (2008). The Dryad Repository Application Profile: Process, Development, and Refinement (Master's Paper). University of North Carolina at Chapel Hill. Retrieved from https://cdr.lib.unc.edu/indexablecontent/uuid:727a2712-f8f0-40a3-8876-e8e16b6d086d.

Coyle, Karen, and Thomas Baker. (2009). Guidelines for Dublin Core Application Profiles. Retrieved from http://dublincore.org/documents/2009/05/18/profile-guidelines/

Creative Commons. (2015). CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. Retrieved from http://creativecommons.org/publicdomain/zero/1.0/

Darwin Core Task Group. (2015). Darwin Core. Retrieved from http://rs.tdwg.org/dwc/

DCMI. (2007). DCMI Abstract Model. Retrieved from http://dublincore.org/documents/abstract-model/

DCMI. (2008). The Singapore Framework for Dublin Core Application Profiles. Retrieved from http://dublincore.org/documents/singapore-framework/

DCMI. (2012). DCMI Metadata Terms. Retrieved from http://dublincore.org/documents/dcmi-terms

DDI. (2009). Data Documentation Initiative. Retrieved from http://www.ddialliance.org/

Dryad. (2013). Metadata Profile. Dryad Wiki. Retrieved from http://wiki.datadryad.org/Metadata_Profile.

Dryad. (2015). The Organization: Overview. Retrieved March 30, 2015, from http://datadryad.org/pages/organization.

Dryad Data Repository. (2015). Retrieved from http://datadryad.org/DSpace. (2015). DSpace. Retrieved from http://http://www.dspace.org/Dube, Jed. (2007). A Metadata Application Profile for the DRIADE Project.

Dube, Jed, Sarah Carrier, and Jane Greenberg. (2007). DRIADE: A data repository for evolutionary biology. Proceedings of the 2007 Conference on Digital Libraries, Vancouver, BC, Canada, ACM Press, pp. 481. doi:10.1145/1255175.1255280

Greenberg, Jane, Sarah Carrier, and Jed Dube. (2007). The DRIADE Project: Phased Application Profile Development in Support of Open Science. International Conference on Dublin Core and Metadata Applications, pp. 35–42.

Greenberg, Jane, Hollie White, Sarah Carrier, and Ryan Scherle. (2009). A Metadata Best Practice for a Scientific Data Repository. Journal of Library Metadata, 9(3), 194-212. doi:10.1080/19386380903405090

Heery, Rachel, and Manjula Patel. (2000). Application profiles: Mixing and matching metadata schemas. Ariadne, 25. Retrieved from http://www.ariadne.ac.uk/issue25/app-profiles/

Idealliance. (2015). PRISM Metadata Initiative. Retrieved from http://www.idealliance.org/specifications/prism-metadata-initiative

Krause, Edward M., Erin Clary, Adrian Ogletree, Jane Greenberg. (2015). Data from: Evolution of an application profile: advancing metadata best practices through the Dryad data repository. Dryad Digital Repository. doi:10.5061/dryad.f0n35

Library of Congress (LoC). (2015). PREMIS Data Dictionary for Preservation Metadata. Retrieved from http://www.loc.gov/standards/premis/

Mannheimer, Sara, Ayoung Yoon, Jane Greenberg, Elena Feinstein, and Ryan Scherle. (2014). A Balancing Act: The Ideal and the Realistic in Developing Dryad's Preservation Policy. First Monday, 19(8). doi:10.5210/fm.v19i8.5415

◉DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

National Information Standards Organization (NISO). (2004). Understanding Metadata. Bethesda, MD: NISO Press. http://www.niso.org/publications/press/UnderstandingMetadata.pdf

NCBI. (2012). Journal Publishing Tag Set. Retrieved from http://dtd.nlm.nih.gov/publishing/

Nevile, Liddy, and Eva Méndez. (2015). Do We Need Application Profiles? Reflections and suggestions from work in DCMI and ISO/IEC. In DC-2015: Proceedings of the International Conference on Dublin Core and Metadata Applications. São Paulo, Brazil, September 1-5, 2015.

Nilsson, Mikael, Thomas Baker, and Pete Johnston. (2008). The Singapore Framework for Dublin Core Application Profiles. Retrieved from http://dublincore.org/documents/singapore-framework/.

Nilsson, Mikael, Alistair J. Miles, Pete Johnston, and Fredrik Enoksson. (2009). Formalizing Dublin Core Application Profiles – Description Set Profiles and Graph Constraints. In Sicilia, Miguel-Angel & Miltiadis D. Lytras (Eds.), Metadata and Semantics, pp. 101–111. Boston, MA: Springer US. Retrieved from http://link.springer.com/10.1007/978-0-387-77745-0_10.

Powell, Andy, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, and Thomas Baker. (2007). DCMI Abstract Model. Retrieved from http://dublincore.org/documents/abstract-model/.

White, Hollie. (2008). Exploring evolutionary biologists' use and perceptions of semantic metadata for data curation. In International Conference on Dublin Core and Metadata Applications 2008, September 22-26, 2008, Berlin, Germany, pp. 202. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/936/932.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

Appendix A: Dryad Metadata Application Profile, Version 3.2

Internal Element Representation (DSpace):				dcterms:contributor.author	
External Element Representation (Metadata Harvesting APIs):				dcterms:creator	
URI:	http://purl.org/dc/terms/creator				
Definition:	Authors on publication / Authors of data submission				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):				dcterms:coverage.spatial	
External Element Representation (Metadata Harvesting APIs):				dcterms:spatial	
URI:	http://purl.org/dc/terms/spatial				
Definition:	Spatial description of the data specified by a geographic description and/or geographic coordinates				
Module(s):	Package & File	**Obligation:**	Optional	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):				dcterms:coverage.temporal	
External Element Representation (Metadata Harvesting APIs):				dcterms:temporal	
URI:	http://purl.org/dc/terms/temporal				
Definition:	Temporal description of the data, as geologic timespan or dates of data collection/research				
Module(s):	Package & File	**Obligation:**	Optional	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):				dcterms:date.accessioned	
External Element Representation (Metadata Harvesting APIs):				dcterms:dateSubmitted	
URI:	http://purl.org/dc/terms/dateSubmitted				
Definition:	Date DSpace takes possession of item after a curator archives the item				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):				dcterms:date.available	
External Element Representation (Metadata Harvesting APIs):				dcterms:available	
URI:	http://purl.org/dc/terms/available				
Definition:	Date and time the package becomes available to the public on DSpace				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):				dcterms:date.blackoutUntil	
External Element Representation (Metadata Harvesting APIs):				N/A; Internal element only	
URI:	URI not assigned				
Definition:	A date after which the dataset will automatically archive itself (move out of publication blackout)				
Module(s):	Package	**Obligation:**	Optional	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):				dcterms:date.embargoedUntil	
External Element Representation (Metadata Harvesting APIs):				N/A; Internal element only	
URI:	URI not assigned				
Definition:	Embargo date - a date after which the dataset will be made public				
Module(s):	File	**Obligation:**	Optional	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:date.issued

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

External Element Representation (Metadata Harvesting APIs):					dcterms:issued
URI:	http://purl.org/dc/terms/issued				
Definition:	Date of journal article publication				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):					dcterms:description
External Element Representation (Metadata Harvesting APIs):					dcterms:description
URI:	http://purl.org/dc/terms/description				
Definition:	Description of entity; In the data package module, refers to abstract of associated scholarly publication				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):					dcterms:description.provenance
External Element Representation (Metadata Harvesting APIs):					dcterms:provenance
URI:	http://purl.org/dc/terms/provenance				
Definition:	Information related to the origin and integrity of the file; history of custody of the item since its creation, including any changes successive custodians made to the item				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):					dcterms:format.extent
External Element Representation (Metadata Harvesting APIs):					dcterms:extent
URI:	http://purl.org/dc/terms/extent				
Definition:	Size of the file (bytes)				
Module(s):	File	**Obligation:**	Required	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):					dcterms:identifier
External Element Representation (Metadata Harvesting APIs):					dcterms:identifier
URI:	http://purl.org/dc/terms/identifier				
Definition:	DOI of the Dryad entity (data package or data file)				
Module(s):	Package & File	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):					dcterms:identifier.citation
External Element Representation (Metadata Harvesting APIs):					dcterms:bibliographicCitation
URI:	http://purl.org/dc/terms/bibliographicCitation				
Definition:	Standard bibliographic citation of the associated scholarly publication				
Module(s):	Package	**Obligation:**	Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):					dcterms:identifier.manuscriptNumber
External Element Representation (Metadata Harvesting APIs):					N/A; Internal element only
URI:	URI not assigned				
Definition:	Manuscript number of associated scholarly publication				
Module(s):	Package	**Obligation:**	Optional	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):					dcterms:identifier.uri
External Element Representation (Metadata Harvesting APIs):					dcterms:identifier
URI:	http://purl.org/dc/terms/identifier				
Definition:	URL which links to the web location of the Dryad entity				

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

Module(s):	Package & File	Obligation:	Required	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:relation.haspart	
External Element Representation (Metadata Harvesting APIs):		dcterms:hasPart	
URI:	http://purl.org/dc/terms/hasPart		
Definition:	Record identifier for associated Dryad data file (doi:###/1 ; doi###/2 ; etc.)		

Module(s):	Package	Obligation:	Required	Cardinality:	Repeatable

Internal Element Representation (DSpace):		dcterms:relation.ispartof	
External Element Representation (Metadata Harvesting APIs):		dcterms:isPartOf	
URI:	http://purl.org/dc/terms/isPartOf		
Definition:	Associated Dryad Data Package Identifier (doi:###) - the "root" doi of the package		

Module(s):	File	Obligation:	Required	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:relation.ispartofseries	
External Element Representation (Metadata Harvesting APIs):		N/A; Internal element only	
URI:	URI not assigned		
Definition:	Series name and number within that series, if available		

Module(s):	Package	Obligation:	Optional	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:rights.uri	
External Element Representation (Metadata Harvesting APIs):		dcterms:rights	
URI:	http://purl.org/dc/terms/rights		
Definition:	Statement regarding the rights held over the resource, e.g. CC0 (Creative, 2015)		

Module(s):	File	Obligation:	Required	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:subject	
External Element Representation (Metadata Harvesting APIs):		dcterms:subject	
URI:	http://purl.org/dc/terms/subject		
Definition:	Keywords associated with the Dryad entity		

Module(s):	Package & File	Obligation:	Optional	Cardinality:	Repeatable

Internal Element Representation (DSpace):		dcterms:title	
External Element Representation (Metadata Harvesting APIs):		dcterms:title	
URI:	http://purl.org/dc/terms/title		
Definition:	Title of entity (article, dataset, package, file, etc.)		

Module(s):	Package & File	Obligation:	Required	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:type	
External Element Representation (Metadata Harvesting APIs):		dcterms:type	
URI:	http://purl.org/dc/terms/type		
Definition:	Entity type: article (package) or dataset (file)		

Module(s):	Package & File	Obligation:	Required	Cardinality:	Non-Repeatable

Internal Element Representation (DSpace):		dcterms:type.embargo	
External Element Representation (Metadata Harvesting APIs):		N/A; Internal element only	

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2015

URI:	URI not assigned			
Definition:	Length of Embargo (none, oneyear, custom)			
Module(s):	File	**Obligation:** Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):	dryad:downloads			
External Element Representation (Metadata Harvesting APIs):	N/A; Internal element only			
URI:	URI not assigned			
Definition:	Number of times the data file has been downloaded			
Module(s):	File	**Obligation:** Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):	dryad.externalIdentifier			
External Element Representation (Metadata Harvesting APIs):	dcterms:identifier			
URI:	http://purl.org/dc/terms/identifier			
Definition:	Unique identifier for related data in Dryad partner repository			
Module(s):	Package	**Obligation:** Optional	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):	dryad:pageviews			
External Element Representation (Metadata Harvesting APIs):	N/A; Internal element only			
URI:	URI not assigned			
Definition:	Number of times the webpage of a data file has been viewed			
Module(s):	File	**Obligation:** Required	**Cardinality:**	Non-Repeatable

Internal Element Representation (DSpace):	dwc:ScientificName			
External Element Representation (Metadata Harvesting APIs):	dwc:scientificName			
URI:	http://rs.tdwg.org/dwc/terms/scientificName			
Definition:	Full name of the lowest level taxon to which the organism has been identified in the most recent accepted determination, specified as precisely as possible (may also specify other levels of biological taxonomy)			
Module(s):	Package & File	**Obligation:** Optional	**Cardinality:**	Repeatable

Internal Element Representation (DSpace):	prism:publicationName			
External Element Representation (Metadata Harvesting APIs):	prism:publicationName			
URI:	http://www.prismstandard.org/specifications/3.0/PRISM_Basic_Metadata_3.0.htm#_Toc336960554			
Definition:	Name of publication associated with an item (i.e. journal name)			
Module(s):	Package	**Obligation:** Required	**Cardinality:**	Non-Repeatable